

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE  
BEFORE THE BOARD OF PATENT APPEALS AND INTERFERENCES

---

In re the Application of

Thorsten H. Brants et al.

Art Unit: 2167

Application No.: 10/626,856

Examiner: Kimberly M. Lovel

Filed: July 25, 20003

Docket No.: A3052-US-NP  
XERZ 201563

Title: SYSTEMS AND METHODS FOR NEW EVENT DETECTION

---

Appeal Brief - Patents  
Commissioner for Patents  
P.O. Box 1450  
Alexandria, VA 22313-1450

**TRANSMITTAL OF  
APPEAL BRIEF UNDER 37 C.F.R. § 41.37**

Dear Sir:

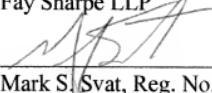
Applicant transmits herewith one copy of the originally signed APPEAL  
BRIEF UNDER 37 C.F.R. § 41.37 for the above-reference patent application.

Payment for the required fee for the filing of this Appeal Brief is being  
authorized via the EFS Web to be charged to Deposit Account No. 24-0037.

Respectfully submitted,  
Fay Sharpe LLP

February 2, 2009

Date

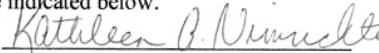
  
Mark S. Svat, Reg. No. 34,261  
The Halle Building – Fifth Floor  
1228 Euclid Avenue  
Cleveland, OH 44115  
Phone: 216-363-9000  
Facsimile: (216) 363-9001  
E-Mail: [msvat@faysharpe.com](mailto:msvat@faysharpe.com)  
Attorneys for Appellant(s)

**CERTIFICATE OF TRANSMISSION**

I hereby certify that this correspondence (and any item referred to herein as being attached or enclosed) is (are) being transmitted to the USPTO by electronic transmission via the EFS Web on the date indicated below.

February 2, 2009

Date

  
Name: Kathleen A. Nimrichter

PATENT APPLICATION

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE  
BEFORE THE HONORABLE BOARD OF PATENT APPEALS AND INTERFERENCES

In re the Application of

Thorsten H. Brants et al.

Application No.: 10/626,856

Examiner: Kimberly M. Lovel

Filed: July 25, 2003

Docket No.: A3052-US-NP  
XERZ 2 01563

Title: SYSTEMS AND METHODS FOR NEW EVENT DETECTION

BRIEF ON APPEAL

Appeal from Group 2167

Mark S. Svat, Esq., Reg. No. 34,261  
FAY SHARPE LLP  
The Halle Building – Fifth Floor  
1228 Euclid Avenue  
Cleveland, OH 44115  
216-363-9000  
Attorneys for Appellants

**CERTIFICATE OF TRANSMISSION**

I hereby certify that this correspondence (and any item referred to herein as being attached or enclosed) is (are) being transmitted to the USPTO by electronic transmission via the EFS Web on the date indicated below.

February 2, 2009  
Date

Kathleen A. Nimrichter  
Name: KATHLEEN A. NIMRICKTER

TABLE OF CONTENTS

	<u>Page</u>
I. REAL PARTY IN INTEREST .....	1
II. RELATED APPEALS AND INTERFERENCES .....	2
III. STATUS OF CLAIMS .....	3
IV. STATUS OF AMENDMENTS .....	4
V. SUMMARY OF CLAIMED SUBJECT MATTER.....	5
VI. GROUNDS OF REJECTION TO BE REVIEWED ON APPEAL .....	10
VII. ARGUMENT .....	11
A. Claims 1, 7, 9-11, and 13-14 (and Remaining Dependent Claims) Would Not Have Been Obvious Over Final Report in View of Relevance Models in View of Brown .....	11
1. Claim 1 .....	11
a. Same Source Characteristic Not Taught by Brown .....	11
b. Relevance Models Does Not Teach Recited Formula.....	17
2. Claim 13 .....	17
a. Same Source Characteristic Not Taught by Brown .....	17
b. Final Report Does Not Teach Recited Formula.....	18
3. Claim 14 .....	18
a. Same Source Characteristic Not Taught by Brown .....	18
b. Final Report Does Not Teach Recited Formula.....	19
4. Claims 2-6, 15, and 37 .....	20
VIII. CLAIMS APPENDIX: .....	22
IX. EVIDENCE APPENDIX .....	27
X. RELATED PROCEEDINGS APPENDIX.....	28

I. REAL PARTY IN INTEREST

The real party in interest for this appeal and the present application is Palo Alto Research Center Inc. (3333 Coyote Hill Rd., Palo Alto, California 94304), by way of an Assignment recorded in the U.S. Patent and Trademark Office at Reel 014325, Frame 0277.

II. RELATED APPEALS AND INTERFERENCES

There are no prior or pending appeals, interferences or judicial proceedings, known to Appellant, Appellant's representative, or the Assignee, that may be related to, or which will directly affect or be directly affected by or have a bearing upon the Board's decision in the pending appeal.

III. STATUS OF CLAIMS

Claims 1-7, 9-11, 13-15, and 37 are on appeal.

Claims 1-7, 9-11, 13-15, and 37 are pending.

Claims 1-7, 9-11, 13-15, and 37 are rejected.

Claims 8, 12, 16-36, and 38-39 are canceled.

IV. STATUS OF AMENDMENTS

An Amendment After Final Rejection was filed on October 6, 2008. By an Advisory Action dated October 20, 2008, it was indicated that the requested amendments had been entered. However, a second Amendment After Final Rejection was filed on November 6, 2008, and by an Advisory Action dated November 17, 2008, it was indicated that the requested amendments had also been entered. Therefore, the claims presented for appeal are those as set forth within Applicants' second Amendment After Final submitted November 6, 2008.

## V. SUMMARY OF CLAIMED SUBJECT MATTER

The invention of claim 1 is directed to a computer-implemented method of detecting new events. Story characteristics are determined based on an average story similarity story characteristic and a same event-same source story characteristic as described with reference to reference number S300 in FIG. 2A, on page 8, lines 25-29. A source-identified story corpus and a source-identified new story are determined, with each story associated with at least one event (FIG. 2A: S400, S500; and page 9, lines 6-20). One or more story-pairs based on the source-identified new-story and each story in the source-identified story corpus, and at least one inter-story similarity metric for the story-pairs are determined (FIG. 2A: S800; and page 10, lines 6-10). The inter-story similarity metrics include one or more story frequency models and story characteristic frequency models combined using terms weights (FIG. 2A: S1000; page 10, lines 16-33). An event frequency is determined based on term  $t$  and  $ROI$  category  $rmax$  from the formula: 
$$ef_{rmax}(t) = \frac{\max}{r \in R} (ef(r, t))$$
, as shown at step S1200 in FIG. 2A, and described on page 11, lines 21-24. One or more adjustments to the inter-story similarity metrics are determined based on one or more story characteristics (FIG. 2A: S1400; page 14, line 19 – page 16, line 11). A new story event indicator is outputted if the event associated with the new story is similar to the events associated with the source-identified story corpus based on the inter-story similarity metrics and the adjustments (FIF. 3: 99, 300; page 21, lines 14-16).

The invention of claim 2 is directed to the method of claim 1, wherein the inter-story similarity metric is dynamically adjusted based on at least one of subtraction and division.

The invention of claim 3 is directed to the method of claim 1, wherein the inter-story similarity metric is either a probability based inter-story similarity metric or a Euclidean based inter-story similarity metric (FIG. 2A: S1600; page 17, lines 1-9).

The invention of claim 4 is directed to the method of claim 3, wherein the probability based inter-story similarity metric is a Hellinger, a Tanimoto, a KL divergence or a clarity distance based metric (FIG. 2A: S1300; page 12, lines 13-16).

The invention of claim 5 is directed to the method of claim 3, wherein the Euclidean based similarity metric is a cosine-distance based metric (FIG. 2A: S1300; page 12, lines 13-16).

The invention of claim 6 is directed to the method of claim 1, wherein the inter-story similarity metrics are determined based on a term frequency-inverse story frequency model (FIG. 2A: S1000; page 10, lines 16-20).

The invention of claim 7 is directed to the method of claim 1, wherein the inter-story similarity metrics are one or more story frequency models or one or more event frequency models, combined using terms weights (page 12, lines 1-6).

The invention of claim 9 is directed to the method of claim 1, where the adjustments based on the story characteristics are applied to the term weights (FIG. 2A: S1500; page 16, lines 6-7).

The invention of claim 10 is directed to the method of claim 1, where the adjustments based on the story characteristics are applied to the inter-story similarity metrics (FIG. 2A: S1500; page 16, lines 6-7).

The invention of claim 11 is directed to the method of claim 1, wherein the inter-story similarity metrics include one or more term frequency-inverse event frequency

models, and where the events are classified based on story labels or a predictive model (FIG. 2A: S1200; page 12, lines 7-11).

The invention of claim 13 is directed to a method of detecting new events. Story characteristics are determined based on an average story similarity story characteristic and a same event-same source story characteristic as described with reference to reference number S300 in FIG. 2A, on page 8, lines 25-29. A source-identified story corpus and a source-identified new story are determined, with each story associated with at least one event (FIG. 2A: S400, S500; and page 9, lines 6-20). One or more story-pairs based on the source-identified new-story and each story in the source-identified story corpus, and at least one inter-story similarity metric for the story-pairs are determined (FIG. 2A: S800; and page 10, lines 6-10). The inter-story similarity metrics include one or more story frequency models and story characteristic frequency models combined using terms weights (FIG. 2A: S1000; page 10, lines 16-33). An inverse event frequency is determined based on term  $t$ , and events  $e$  and  $r_{\max}$  in the

set of  $ROI$  categories from the formula:  $IEF(t) = \log \left[ \frac{N_{e,r_{\max}}}{ef_{r_{\max}}(t)} \right]$ , as shown at step S1200

in FIG. 2A, and described on page 11, line 27 – page 12, line 1. One or more adjustments to the inter-story similarity metrics are determined based on one or more story characteristics (FIG. 2A: S1400; page 14, line 19 – page 16, line 11). A new story event indicator is outputted if the event associated with the new story is similar to the events associated with the source-identified story corpus based on the inter-story similarity metrics and the adjustments (FIF. 3: 99, 300; page 21, lines 14-16).

The invention of claim 14 is directed to a method of detecting new events. Story characteristics are determined based on an average story similarity story characteristic

and a same event-same source story characteristic as described with reference to reference number S300 in FIG. 2A, on page 8, lines 25-29. A source-identified story corpus and a source-identified new story are determined, with each story associated with at least one event (FIG. 2A: S400, S500; and page 9, lines 6-20). One or more story-pairs based on the source-identified new-story and each story in the source-identified story corpus, and at least one inter-story similarity metric for the story-pairs are determined (FIG. 2A: S800; and page 10, lines 6-10). The inter-story similarity metrics include one or more story frequency models and story characteristic frequency models combined using terms weights (FIG. 2A: S1000; page 10, lines 16-33). An inverse event frequency is determined based on term  $t$ , categories  $e$ ,  $r$  and  $r_{max}$  in the set of  $ROI$  categories and  $P(r)$ , the probability of  $ROI/r$  from the formula:

$$IEF'(t) = \sum_{r \in R} P(r) \log \left[ \frac{N_{e,r}}{ef(r,t)} \right], \text{ as shown at step S1200 in FIG. 2A, and described on}$$

page 12, lines 1-6. One or more adjustments to the inter-story similarity metrics are determined based on one or more story characteristics (FIG. 2A: S1400; page 14, line 19 – page 16, line 11). A new story event indicator is outputted if the event associated with the new story is similar to the events associated with the source-identified story corpus based on the inter-story similarity metrics and the adjustments (FIF. 3: 99, 300; page 21, lines 14-16).

The invention of claim 15 is directed to the method of claim 1, further including the step of determining a subset of stories from the source-identified story corpus and the source-identified new story based on one or more story characteristics.

The invention of claim 37 is directed to the method of claim 1, in which the new event indicator is displayed on a visual, audio or tactile output device (FIG. 300: 300, 400; page 7, lines 27-29; and page 21, lines 14-17).

VI. GROUNDS OF REJECTION TO BE REVIEWED ON APPEAL

The following grounds of rejection are presented for review:

- 1) Claims 1, 7, 9-11, and 13-14 are rejected as having been obvious under 35 U.S.C. §103(a) over the article "Topic Detection and Tracking Pilot Study Final Report" by Allan et al. (hereinafter Final Report) in view of the article "Relevance Models for Topic Detection and Tracking" by Lavrenko et al. (hereinafter Relevance Models) in view of the article "Dynamic Stopwording for Story Link Detection" by Brown. (hereinafter Brown).
- 2) Whether remaining dependent claims 2-6, 15, and 37 are patentable over the references.

VII. ARGUMENT

A. Claims 1, 7, 9-11, and 13-14 (and Remaining Dependent Claims) Would Not Have Been Obvious Over Final Report in View of Relevance Models in View of Brown

1. Claim 1

a. Same Source Characteristic Not Taught by Brown

With reference to claim 1, turning to section 5, (page 5) of the final Office Action of July 11, 2008, it is there stated that the combination of Final Report and Relevance Model discloses the concepts of determining at least one story characteristic based on an average similarity story characteristic, a story corpus and a new story. It then notes that the combination fails to explicitly disclose the further limitations of determining at least one story characteristic based on an average similarity story characteristic and a same event-same source story characteristic, a source-identified story corpus and source-identified new story. The Office Action then argues, however, that "Brown discloses topic detection and tracking, including the further limitations of determining at least one story characteristic based on an average similarity story characteristic [determining whether two news stories discuss the same subject] and a same event-same source story characteristic [a dual threshold is used to determine whether the computed cosine similarity indicates linkage between the two stories; one threshold is used when the two documents originate from the same type of source, and the other threshold is used for documents from different sources]." The Office Action therein makes reference to Section 1: Introduction and Section 2: System Description, 1st Paragraph.

Applicants argued in the second Amendment After Final submitted November 6, 2008, that the teachings of Brown regarding sources are different than the same-source limitations recited in claim 1 and other claims of the present application. Applicants further argued that Brown teaches against the same event-same source characteristic as recited in the claim. Applicants continue to maintain these arguments which are reiterated in the following paragraphs.

Before discussing the teachings of Brown regarding sources, Applicants first briefly discussed the same source characteristic described in the present application and recited in the claim as follows. When discussing similarity adjustments based on event and source-pair information, with specific reference to the source information, paragraph 59 of the original application makes the following statement: "That is, stories describing the same event and originating from the same source tend to use similar language. Thus, two stories originating from the "CNN" source tend to use similar language when describing the same event. Same event-same source adjustments are used to dynamically and selectively change the importance of terms used to determine a similarity metric to compensate for these shared terms and vocabularies" (underlining added for emphasis). This makes it clear that the source information of which the present application makes use of is essentially an authorship type of source, i.e., which organization produces the information. And, the stated reason for doing so is because it is expected that stories emanating from the same source are authored by persons who share a common vocabulary. This is made even more clear in paragraph 81 where the present application notes that two "stories describing the same rule of interpretation category are likely to share vocabulary. Two stories originating from the same source and describing the same event are even more likely to share vocabulary. In various

exemplary embodiments according to this invention, adjustments are determined based on the average of the similarity metrics for each set of same event and same source stories. In this way, the effect of shared vocabulary terms is reduced."

An example of the effect of shared vocabulary of stories originating from the same source, but for different events is offered in paragraph 82: "For example, a first Washington Post Business Section story describes the introduction of company ABC's newest medical product. A second Washington Post Business Section story describes the introduction of company XYZ's newest medical product. These stories are likely to share a large number of vocabulary terms making differentiation difficult. However, if the two Washington Post sourced stories are determined to be from the same ROI, same event-same source adjustments are dynamically and selectively applied to differentiate the stories by reducing the effect of the shared vocabulary terms."

Further description of the same source information which reinforces the above-provided description can be found in paragraphs 90-91, 95 and 97 of the present application. However, it is clear from the description in the application, that in reference to source-identified stories, it is the source that produces the story which is utilized in determining the adjustments to the similarity metrics as recited in the claims.

Also in the second Amendment After Final, Applicants argued that Brown teaches against the use of source identification as recited in claim 1 of the present application. Section 2, SYSTEM DESCRIPTION, first paragraph of the Brown reference, offers the following regarding dual thresholds: "A dual threshold is used to determine whether the computed cosine similarity indicates a linkage between the two stories; one threshold is used when the two documents originate from the same type of source, and the other (typically lower) threshold is used for documents from different sources. For

the 2001 evaluation, all Mandarin-language sources were grouped together, as were all English-language sources (i.e. a news story from Mandarin-language Voice of America would be treated as coming from the same source as a news story from Xinhua newswire)".

The above-quoted parenthetical expressions bear repeating here: "a news story from Mandarin-language Voice of America would be treated as coming from the same source as a news story from Xinhua newswire" (underlining added). Applicants note here that the above-mentioned Voice of America and the Xinhua newswire would clearly be considered as different sources according to the present application, regardless of the story subject matter. Contrariwise, Brown teaches against the present application, and treats these obviously different sources as the same source simply because the documents share a common language, i.e. Mandarin or English. Clearly, Brown is concerned with the type of source, e.g., Mandarin-language vs. English-language, rather than the actual source of the documents.

Further, use of the phrase "same source" in the specification and claim limitations of the present application is consistent with respect to its intended meaning, i.e., the actual source of the information, or in other words, the origin of the information. This is consistent with customary and normal usage and the dictionary definition of the term "source", i.e., a point of origin or procurement, one that initiates, author, one that supplies information, and a firsthand document or primary reference work. Applicants submit that the Office Action and the Advisory Action are incorrectly interpreting the phrase "same source" too broadly to include multiple sources having a similar characteristic, e.g., Mandarin language. Applicants further submit that the Brown reference, as described above, is concerned with characteristics of the source, and

teaches against use of the origin of information as evidenced by the treating of multiple sources as the same source.

However, in the Advisory Action mailed November 17, 2008 in response to the second Amendment After Final, the examiner disagreed that Brown fails to meet the requirements of the recited claim limitations. The Examiner agreed with Applicants' description of what is meant by the phrase same source in the specification, and also agreed that the description was supported by examples from the specification. However, the Examiner asserted that the examples in the specification fail to explicitly limit what types of sources meet the claim limitation, and consequently noted that, although the claims are interpreted in light of the specification, limitations from the specification are not read into the claims. The Examiner argued that, in "Brown, the example sorts documents based on source, where the sources are represented by type of source (i.e., the language of the source). Brown considers two documents to be from the same source as long as they originate from the same type of source. Since, the claim language fails to explicitly limit what is meant by a same source, Brown is considered to meet the requirements of the claimed same source."

Applicants respectfully disagree with the Examiner's position that the examples in the specification fail to explicitly limit what types of sources meet the claim limitation. The specification is very consistent in its use of the phrase "same source" as discussed above with reference to paragraphs 58 and 81 of the present application. It is clear that the specification is describing only and authorship type of source, i.e., which organization produces the information. This limitation is reinforced by the association of the "same event" characteristic in both the specification and claim recitations. The specification and claims are clearly limiting the "same source" characteristic to the

above-described types of sources. Applying a broader interpretation of the recited phrase such as, e.g., sources of the same language would destroy claimed functionality which is based, as described in the specification, on the identification actual sources of the information, not on a broad characterization of the type of source.

For at least the above-stated reasons, Applicants submit that the specification clearly defines the phrases "same source" and "source-identified" as recited in the claims. Therefore, Applicants also respectfully submit that the recitation of these phrases in claim 1 of the present application limits the interpretation of the recited phrases, and that the Examiner has erred in interpreting the phrase more broadly.

Based on the preceding discussion, Applicants respectfully submit that the Brown reference teaches considering different document sources as the same source based on other characteristics of the source (e.g., language). Thus, a modification of the Brown reference is necessary to at least potentially meet the claimed invention of the present application. Because this modification would obviously destroy the function of the Brown reference, one of ordinary skill in the art would not find reason to make the modification. In other words, the Brown reference is not properly combinable with present application because the intended function of the Brown reference would be destroyed by the necessary modification.

For at least the above-described reasons, Applicants submit that the recited "same source" limitations of claim 1 render the claim patentably distinct over the references.

b. Relevance Models Does Not Teach Recited Formula

Claim 1, as amended in the second Amendment After Final, recites a formula for an event frequency. Applicants submit that Relevance Models does not describe a formula which teaches or suggests the formula recited in claim 1. The Examiner, however, disagrees that Relevance Models fails to teach the formula, and considers the formula of Relevance Model to be equivalent to that of the claimed invention.

Applicants respectfully submit that the Examiner is interpreting the equations described in Relevance Models too broadly. The formula recited in claim 1 of the present application for the event frequency is described in paragraph 42 of the specification, and Applicant submits that the Examiner has not shown where Relevance Models describes a similar formula. In particular, the Examiner has not shown where Relevance Models describes a formula for event frequency.

For at least the above-described reasons, Applicants submit that the recited formula for an event frequency, further renders claim 1 patentably distinct over the references.

2. Claim 13

a. Same Source Characteristic Not Taught by Brown

Claim 13, as amended in the second After Final Amendment, recites the same limitations regarding the "same source" limitation as recited in claim 1 as discussed above. Applicants submit, therefore, that the arguments presented above with reference to the same source characteristic apply as well to claim 13.

For at least the above-described reasons, Applicants submit that the recited "same source" limitations of claim 13 render the claim patentably distinct over the references.

b. Final Report Does Not Teach Recited Formula

Claim 13, as amended in the second Amendment After Final, recites a formula for an inverse event frequency. Applicants submit that Final Report does not describe a formula which teaches or suggests the formula recited in claim 13. The Examiner, however, disagrees that Final Report fails to teach the formula, and considers the formula of Final Report to be equivalent to that of the claimed invention.

Applicants respectfully submit that the Examiner is interpreting the equations described in Final Report too broadly. The formula recited in claim 13 of the present application for the inverse event frequency is described in paragraph 43 of the specification, and Applicant submits that the Examiner has not shown where Final Report describes a similar formula. For example, the equation set forth in Final Report for *IDF* is a parametric equation, comprising a function of *x* (current point) and *t* (term), whereas the formula for *IEF* recited in claim 13 is a function only of *t*.

For at least the above-described reasons, Applicants submit that the recited formula for an inverse event frequency, further renders claim 13 patentably distinct over the references.

3. Claim 14

a. Same Source Characteristic Not Taught by Brown

Claim 14, as amended in the second After Final Amendment, recites the same limitations regarding the "same source" limitation as recited in claim 1 as discussed

above. Applicants submit, therefore, that the arguments presented above with reference to the same source characteristic apply as well to claim 14.

For at least the above-described reasons, Applicants submit that the recited "same source" limitations of claim 14 render the claim patentably distinct over the references.

b. Final Report Does Not Teach Recited Formula

Claim 14, as amended in the second Amendment After Final, recites a formula for an inverse event frequency. Applicants submit that Final Report does not describe a formula which teaches or suggests the formula recited in claim 14. The Examiner, however, disagrees that Final Report fails to teach the formula, and considers the formula of Final Report to be equivalent to that of the claimed invention.

Applicants respectfully submit that the Examiner is interpreting the equations described in Final Report too broadly. The formula recited in claim 14 of the present application for the inverse event frequency is described in paragraph 43 of the specification, and Applicant submits that the Examiner has not shown where Final Report describes a similar formula. For example, the equation set forth in Final Report for *IDF* is a parametric equation, comprising a function of *x* (current point) and *t* (term), whereas the formula for *IEF* recited in claim 13 is a function only of *t*.

For at least the above-described reasons, Applicants submit that the recited formula for an inverse event frequency, further renders claim 14 patentably distinct over the references.

4. Claims 2-6, 15, and 37

Each of claims 2-6, 15, and 37 depend from and further define distinguished independent claim 1.

CONCLUSION

For all of the reasons discussed above, it is respectfully submitted that the rejections are in error and that independent claims 1, 13, and 14 are in condition for allowance. Applicants submit also that each of the remaining dependent claims 2-7, 9-11, 15, and 37, by reason of dependence from their respective base claims, are also in condition for allowance. For all of the above reasons, Appellants respectfully request this Honorable Board to reverse the rejections of claims 1-7, 9-11, 13-15, and 37.

Respectfully submitted,

  
\_\_\_\_\_  
Mark S. Svat

Registration No. 34,261

: /

FAY SHARPE LLP  
The Halle Building – 5<sup>th</sup> Floor  
1228 Euclid Avenue  
Cleveland, Ohio 44115  
Telephone: (216) 363-9000

Filed: 2/2/09

## APPENDICES

### VIII. CLAIMS APPENDIX:

Claims involved in the Appeal are as follows:

1. A computer-implemented method of detecting new events comprising the steps of:

determining at least one story characteristic based on an average story similarity story characteristic and a same event-same source story characteristic;

determining a source-identified story corpus, each story associated with at least one event;

determining a source-identified new story associated with at least one event;

determining story-pairs based on the source-identified new-story and each story in the source-identified story corpus;

determining at least one inter-story similarity metric for the story-pairs; wherein the inter-story similarity metrics are comprised of at least one story frequency model; and at least one story characteristic frequency model combined using terms weights; and wherein an event frequency is determined based on term  $t$  and  $ROI$  category  $r_{max}$

from the formula:  $ef_{r_{max}}(t) = \max_{r \in R} (ef(r, t))$ ;

determining at least one adjustment to the inter-story similarity metrics based on at least one story characteristic; and

outputting a new story event indicator if the event associated with the new story is similar to the events associated with the source-identified story corpus based on the inter-story similarity metrics and the adjustments.

2. The method of claim 1, wherein the inter-story similarity metric is dynamically adjusted based on at least one of subtraction and division.
3. The method of claim 1, wherein the inter-story similarity metric is at least one of a probability based inter-story similarity metric and a Euclidean based inter-story similarity metric.
4. The method of claim 3, wherein the probability based inter-story similarity metric is at least one of a Hellinger, a Tanimoto, a KL divergence and a clarity distance based metric.
5. The method of claim 3, wherein the Euclidean based similarity metric is a cosine-distance based metric.
6. The method of claim 1, wherein the inter-story similarity metrics are determined based on a term frequency-inverse story frequency model.
7. The method of claim 1, wherein the inter-story similarity metrics are comprised of: at least one story frequency model; and at least one event frequency model combined using terms weights.
9. The method of claim 1, where the adjustments based on the story characteristics are applied to the term weights.
10. The method of claim 1, where the adjustments based on the story characteristics are applied to the inter-story similarity metrics.

11. The method of claim 1, wherein the inter-story similarity metrics are comprised of at least one term frequency-inverse event frequency model and where the events are classified based on at least one of: story labels and a predictive model.

13. A computer-implemented method of detecting new events comprising the steps of:

determining at least one story characteristic based on an average story similarity story characteristic and a same event-same source story characteristic;

determining a source-identified story corpus, each story associated with at least one event;

determining a source-identified new story associated with at least one event;

determining story-pairs based on the source-identified new-story and each story in the source-identified story corpus;

determining at least one inter-story similarity metric for the story-pairs; wherein the inter-story similarity metrics are comprised of at least one story frequency model; and at least one story characteristic frequency model combined using terms weights; and wherein an inverse event frequency is determined based on term  $t$ , and events  $e$

and  $r_{\max}$  in the set of  $ROI$  categories from the formula:  $IEF(t) = \log \left[ \frac{N_{e,r_{\max}}}{ef_{r_{\max}}(t)} \right]$ ;

determining at least one adjustment to the inter-story similarity metrics based on at least one story characteristic; and

outputting a new story event indicator if the event associated with the new story is similar to the events associated with the source-identified story corpus based on the inter-story similarity metrics and the adjustments.

14. A computer-implemented method of detecting new events comprising the steps of:

determining at least one story characteristic based on an average story similarity story characteristic and a same event-same source story characteristic;

determining a source-identified story corpus, each story associated with at least one event;

determining a source-identified new story associated with at least one event;

determining story-pairs based on the source-identified new-story and each story in the source-identified story corpus;

determining at least one inter-story similarity metric for the story-pairs; wherein the inter-story similarity metrics are comprised of at least one story frequency model; and at least one story characteristic frequency model combined using terms weights; and wherein an inverse event frequency is determined based on term  $t$ , categories  $e$ ,  $r$  and  $r_{max}$  in the set of  $ROI$  categories and  $P(r)$ , the probability of  $ROI/r$  from the formula:

$$IEF'(t) = \sum_{r \in R} P(r) \log \left[ \frac{N_{e,r}}{ef(r,t)} \right];$$

determining at least one adjustment to the inter-story similarity metrics based on at least one story characteristic; and

outputting a new story event indicator if the event associated with the new story is similar to the events associated with the source-identified story corpus based on the inter-story similarity metrics and the adjustments.

15. The method of claim 1 further comprising the step of determining a subset of stories from the source-identified story corpus and the source-identified new story based on at least one story characteristic.

37. The computer-implemented method of claim 1, in which the new event indicator is displayed on at least one of a visual, audio or tactile output device.

IX. EVIDENCE APPENDIX

NONE

X. RELATED PROCEEDINGS APPENDIX

NONE